Title: Probability, Statistics, and M&Ms

Brief Overview:

Mars, Inc. publishes data concerning the frequency of the colors of M&Ms contained in a package of the candy. In this experiment, students will gather their own data on the frequency of the colors, and compare their values with the expected values according to Mars, Inc.'s data. This unit is designed to enhance students' understanding of basic probabilistic and statistical concepts, such as expected value. It also can be used for more advanced work in a statistics course by having students perform a Chi-squared goodness-of-fit test.

Links to NCTM 2000 Standards:

- **Mathematics as Problem Solving**
  Students will demonstrate their ability to devise a hypothesis, collect and compile data, and use statistical concepts to compare their findings with their hypotheses.

- **Mathematics as Reasoning and Proof**
  Students will demonstrate their ability to reason through the development and testing of their hypotheses, and the justification of their conclusions

- **Mathematics as Communication**
  Students will demonstrate their ability to communicate mathematically through the use of data and the statistical analysis of the data to make a valid conclusion about their hypotheses.

- **Mathematics as Connections**
  Students will recognize the connection among probability, the development of hypotheses, and statistical tests.

- **Mathematics as Representation**
  Students will use M&M candy and the probability of the colors as a model in their Chi-squared goodness-of-fit test.

Links to Maryland High School Mathematics Core Learning Goals:

- **Functions and Algebra**
  Students will demonstrate their ability to use algebraic functions to carry out the calculations during this statistical test.

- **Statistics**
  Students will demonstrate their ability to collect, organize, analyze, and display the data collected throughout the experiment. They will also demonstrate their understanding of the process of hypothesis testing through the analysis of the data.

- **Probability**
  Students will demonstrate their understanding of probability through the discussion of probability distributions and how they relate to the test being performed with respect to the null and alternative hypotheses.

Grade/Level:

This activity is appropriate for grades 7-12.

Duration/Length:

This activity should take approximately 1-2 hours/class periods, depending upon backgrounds and ability levels of the students.

Prerequisite Knowledge:

Students should have working knowledge of the following skills:

- The basic axioms of probability, i.e., all probabilities are between 0 and 1, and the probabilities of all events in the sample space sum to 1
- Expected value (students should be able calculate the frequencies they would expect to see if Mars, Inc.'s data is accurate.)
- The purpose and procedures of hypothesis generation and testing (for more advanced students)
- Chi-squared tests, including how to read a Chi-squared table to obtain a critical value and a p-value for the test (for more advanced students)

Student Outcomes:

Students will:

- gather their own data regarding the frequency of the colors of M&Ms in a package by counting the number of M&Ms in their pack of candy.
- compute the expected number of each color of M&M, given the number of M&Ms in their pack of candy.
- compare their data with the expected values to determine if the data published by Mars, Inc. is consistent with their findings.
- use their data and the expected frequencies to compute a Chi-squared value and subsequently conduct a Chi-squared goodness-of-fit test (for more advanced students).

Materials/Resources/Printed Materials:

- M&Ms (approximately 50-100 for each student)
- Data sheet with published frequencies of colors of M&Ms
- Chi-squared table

Development/Procedures:

Students will begin by sorting their M&Ms by color, counting the number of each color as well as the total number of M&Ms, and recording this on the data sheet. (Note that this experiment can be done individually or as a class, with the teacher collecting and pooling the data from each of the students.) Using the number of M&Ms in the sample and the frequencies published by Mars, Inc., students will use the formula on the data sheet to record the expected number of each color of M&M.

At this point, the exercise can take one of two directions, at the teacher's discretion. For students not familiar with hypothesis testing, the teacher should lead a discussion in which students compare their data with the expected values, commenting on any unusually large discrepancies.

For more advanced students, the teacher should have the class conduct a hypothesis test, with the null hypothesis being that the published data is correct, and the alternative hypothesis being that at least one of the published frequencies is incorrect. Students should choose an appropriate alpha level for the test, typically alpha=0.05. Now, using the formula on the data sheet, students should calculate their Chi-squared value for the test. Having done this, students should use the Chi-squared table to look up a critical value for their test and given alpha level. (Note that for this test, the number of degrees of freedom is the number of colors minus one, or df=5.) Students should now compare their Chi-squared value with the critical value and draw a conclusion by either retaining or rejecting the null hypothesis. Students may also use their Chi-squared value and the Chi-squared table to look up a p-value for their test. The teacher should then lead a discussion concerning the students' results.

Assessment:

Students should present the information in some form of a report. Even if the analysis is done as a class, each individual student should be able to explain the development and results of this statistical test. The students will be scored according to the rubric included.

Extension/Follow Up:

As an extension and an opportunity for further assessment, students could conduct another test following this same procedure. Some possibilities include several hundred rolls of a die, or to show that the number of chocolate chips in a chocolate chip cookie is normally distributed. Students should be encouraged to come up with additional tests they would like to conduct.

Authors:

Rebecca Kenzakowski
Sudbrook Magnet Middle School
Baltimore County, MD

Christopher Wagner
Johns Hopkins University
Baltimore, MD

References:

Chi-Squared table taken from Mendenhall, W. and Beaver, R. Introduction to Probability and Statistics. Duxbury Press, Belmont, California, 1994.

Statistical Analysis Scoring Rubric

3 Hypothesis, collection, and analysis of data show complete understanding of probability, statistics, and the Chi-squared goodness-of-fit test.

2 Hypothesis, collection, and analysis of data show some understanding of probability, statistics, and the Chi-squared goodness-of-fit test.

1 Hypothesis, collection, and analysis of data show little or no understanding of probability, statistics, and the Chi-squared goodness-of-fit test.

M&M Tally

Orange	Red	Brown	Blue	Green	Yellow	TOTAL

Totals:

M&M Analysis

Color	Expected $[E(n_i)]$	Actual (n_i)	Actual – Expected $[n_i - E(n_i)]$	$[n_i - E(n_i)]^2$	$\frac{[n_i - E(n_i)]^2}{E(n_I)}$
Orange					
Red					
Brown					
Blue					
Green					
Yellow					

Calculations:

$[E(n_i)]$ = % probability x Total number M&Ms

(n_i) = actual count / Total number M&Ms

$$x^2 = \sum_{1}^{6} \frac{[n_i - E(n_i)]^2}{E(n_I)}$$

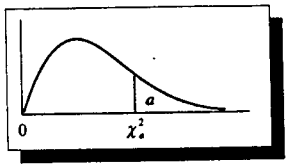

TABLE 5
Critical Values of Chi-Square

d.f.	$\chi^2_{0.995}$	$\chi^2_{0.990}$	$\chi^2_{0.975}$	$\chi^2_{0.950}$	$\chi^2_{0.900}$
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	1.145476	1.61031
6	0.675727	0.872085	1.237347	0.63539	2.20413
7	0.989265	1.239043	1.68987	2.16735	2.83311
8	1.344419	1.646482	2.17973	2.73264	3.48954
9	1.734926	2.087912	2.70039	3.32511	4.16816
10	2.15585	2.55821	3.24697	3.94030	4.86518
11	2.60321	3.05347	3.81575	4.57481	5.57779
12	3.07382	3.57056	4.40379	5.22603	6.30380
13	3.56503	4.10691	5.00874	5.89186	7.04150
14	4.07468	4.66043	5.62872	6.57063	7.78953
15	4.60094	5.22935	6.26214	7.26094	8.54675
16	5.14224	5.81221	6.90766	7.96164	9.31223
17	5.69724	6.40776	7.56418	8.67176	10.0852
18	6.26481	7.01491	8.23075	9.39046	10.8649
19	6.84398	7.63273	8.90655	10.1170	11.6509
20	7.43386	8.26040	9.59083	10.8508	12.4426
21	8.03366	8.89720	10.28293	11.5913	13.2396
22	8.64272	9.54249	10.9823	12.3380	14.0415
23	9.26042	10.19567	11.6885	13.0905	14.8479
24	9.88623	10.8564	12.4011	13.8484	15.6587
25	10.5197	11.5240	13.1197	14.6114	16.4734
26	11.1603	12.1981	13.8439	15.3791	17.2919
27	11.8076	12.8786	14.5733	16.1513	18.1138
28	12.4613	13.5648	15.3079	16.9279	18.9392
29	13.1211	14.2565	16.0471	17.7083	19.7677
30	13.7867	14.9535	16.7908	18.4926	20.5992
40	20.7065	22.1643	24.4331	26.5093	29.0505
50	27.9907	29.7067	32.3574	34.7642	37.6886
60	35.5346	37.4848	40.4817	43.1879	46.4589
70	43.2752	45.4418	48.7576	51.7393	55.3290
80	51.1720	53.5400	57.1532	60.3915	64.2778
90	59.1963	61.7541	65.6466	69.1260	73.2912
100	67.3276	70.0648	74.2219	77.9295	82.3581

Source: From "Tables of the Percentage Points of the χ^2-Distribution," *Biometrika Tables for Statisticians* 1, 3d ed. (1966). Reproduced by permission of the *Biometrika* Trustees.

TABLE 5
(Continued)

$\chi^2_{0.100}$	$\chi^2_{0.050}$	$\chi^2_{0.025}$	$\chi^2_{0.010}$	$\chi^2_{0.005}$	d.f.
2.70554	3.84146	5.02389	6.63490	7.87944	1
4.60517	5.99147	7.37776	9.21034	10.5966	2
6.25139	7.81473	9.34840	11.3449	12.8381	3
7.77944	9.48773	11.1433	13.2767	14.8602	4
9.23635	11.0705	12.8325	15.0863	16.7496	5
10.6446	12.5916	14.4494	16.8119	18.5476	6
12.0170	14.0671	16.0128	18.4753	20.2777	7
13.3616	15.5073	17.5346	20.0902	21.9550	8
14.6837	16.9190	19.0228	21.6660	23.5893	9
15.9871	18.3070	20.4831	23.2093	25.1882	10
17.2750	19.6751	21.9200	24.7250	26.7569	11
18.5494	21.0261	23.3367	26.2170	28.2995	12
19.8119	22.3621	24.7356	27.6883	29.8194	13
21.0642	23.6848	26.1190	29.1413	31.3193	14
22.3072	24.9958	27.4884	30.5779	32.8013	15
23.5418	26.2962	28.8485	31.9999	34.2672	16
24.7690	27.8571	30.1910	33.4087	35.7185	17
25.9894	28.8693	31.5264	34.8053	37.1564	18
27.2036	30.1435	32.8523	36.1908	38.5822	19
28.4120	31.4104	34.1696	37.5662	39.9968	20
29.6151	32.6705	35.4789	38.9321	41.4010	21
30.8133	33.9244	36.7807	40.2894	42.7956	22
32.0069	35.1725	38.0757	41.6384	44.1813	23
33.1963	36.4151	39.3641	42.9798	45.5585	24
34.3816	37.6525	40.6465	44.3141	46.9278	25
35.5631	38.8852	41.9232	45.6417	48.2899	26
36.7412	40.1133	43.1944	46.9630	49.6449	27
37.9159	41.3372	44.4607	48.2782	50.9933	28
39.0875	42.5569	45.7222	49.5879	52.3356	29
40.2560	43.7729	46.9792	50.8922	53.6720	30
51.8050	55.7585	59.3417	63.6907	66.7659	40
63.1671	67.5048	71.4202	76.1539	79.4900	50
74.3970	79.0819	83.2976	88.3794	91.9517	60
85.5271	90.5312	95.0231	100.425	104.215	70
96.5782	101.879	106.629	112.329	116.321	80
107.565	113.145	118.136	124.116	128.299	90
118.498	124.342	129.561	135.807	140.169	100